Generating Single Granularity-Based Fuzzy Classification Rules for Multiobjective Genetic Fuzzy Rule Selection

Rafael Alcalá Yusuke Nojima Francisco Herrera Hisao Ishibuchi

Abstract— **Recently, multiobjective evolutionary algorithms have been applied to improve the difficult tradeoff between interpretability and accuracy of fuzzy rule-based systems. It is known that both requirements are usually contradictory, however, these kinds of algorithms can obtain a set of solutions with different trade-offs.**

The application of multiobjective evolutionary algorithms to fuzzy rule-based systems is often referred to as multiobjective genetic fuzzy systems. The first study on multiobjective genetic fuzzy systems was multiobjective genetic fuzzy rule selection in order to simultaneously achieve accuracy maximization and complexity minimization. This approach is based on the generation of a set of candidate fuzzy classification rules by considering a previously fixed granularity or multiple fuzzy partitions with different granularities for each attribute. Then, a multiobjective evolutionary optimization algorithm is applied to perform fuzzy rule selection. Although the multiple granularity approach is one of the most promising approaches, its interpretability loss has often been pointed out.

In this work, we propose a mechanism to generate single granularity-based fuzzy classification rules for multiobjective genetic fuzzy rule selection. This mechanism is able to specify appropriate single granularities for fuzzy rule extraction before performing multiobjective genetic fuzzy rule selection. The results show that the performance of the obtained classifiers can be even improved by avoiding multiple granularities, which increases the linguistic interpretability of the obtained models.

I. Introduction

Many automatic techniques have been proposed in the literature to extract a proper set of fuzzy rules from numerical data. Most of these techniques usually try to improve the performance associated to the prediction error without paying a special attention to the system interpretability, an essential aspect of fuzzy rule-based systems. In the last years, the problem of finding the right trade-off between interpretability and accuracy, in spite of the original nature of fuzzy logic, has arisen a growing interest in methods which take both aspects into account [4]. Of course, the ideal thing would be to satisfy both criteria to a high degree, but since they are contradictory issues generally it is not possible.

Evolutionary Multiobjective Optimization (EMO) algorithms [6], [7] generate a family of equally valid solutions, where each solution tends to satisfy a criterion to a higher extent than another. For this reason, EMO algorithms have been also applied to improve the accuracy-interpretability trade-off of fuzzy rule-based systems [5], [9], [10], [11], [12], where each solution in the Pareto front represents a different trade-off between interpretability and accuracy (typically measured as complexity and prediction error).

Some of the most recognized works [10], [11] were devoted to the application of EMO algorithms to perform a genetic fuzzy rule selection on an initial set of classification rules involving *"don't care"* conditions and considering two different objectives, classification accuracy and the number of rules. Then, a third objective was also included in order to minimize the length of the rules in [12].

In genetic fuzzy rule selection, a previously fixed granularity [10], [11] or multiple granularities [12] of triangular fuzzy membership functions have been used for the design of fuzzy classifiers, even regression models [1], since an appropriate granularity for each attribute is not known beforehand. By using multiple granularities, the number of fuzzy rules can be successfully reduced in a model. Although the multiple granularity approach is one of the most promising approaches, its interpretability loss has often been pointed out.

To solve the above-mentioned problem, in this work, we propose a single granularity specification for multiobjective genetic fuzzy rule selection. Multiobjective genetic fuzzy rule selection is the following two-step method. In the first phase, a prespecified number of promising fuzzy rules are generated by a heuristic procedure. In the second phase, a multiobjective genetic algorithm is used to select a small number of fuzzy rules from the extracted ones in the first phase. A single granularity specification is an additional process before the second phase. After extracting a prespecified number of fuzzy rules with multiple granularities, a single granularity is specified for each attribute individually according to the frequency of employed partitions and the importance of the multiple granularity-based extracted rules. Then a prespecified number of fuzzy rules are extracted again based on the specified granularity for each attribute.

Following the same main idea, four different mechanisms have been proposed and compared with the original approach on a set of 15 well-known datasets. As well as the interpretability improvement that the use of a single granularity involves, the results show that the performance of the obtained classifiers can be even improved by avoiding multiple granularities.

This contribution is arranged as follows. Next section introduces fuzzy rule-based classifiers by describing the rule structure and inference used in this paper. Section III presents the algorithm proposed to generate single granularity-based

Rafael Alcalá and Francisco Herrera are with the Department of Computer Science and A.I., University of Granada, E-18071 Granada, Spain (email: {alcala, herrera}@decsai.ugr.es).

Yusuke Nojima and Hisao Ishibuchi are with the Department of Computer Science and Intelligent Systems, Osaka Prefecture University, 1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531, Japan (email: {nojima, hisaoi}@cs.osakafu-u.ac.jp).

Supported by the Spanish CICYT Project TIN2005-08386-C05-01 and the Japanese Grant-in-Aid for Young Scientists (B): KAKENHI (18700228).

978-1-4244-3597-5/09/$25.00 ©2009 IEEE

fuzzy classification rules for multiobjective genetic fuzzy rule selection. Section IV shows an experimental study of this method on a set of 15 well-known datasets. Finally, Section V points out some conclusions.

II. Preliminaries: Fuzzy Rule-Based Classifiers Structure and Inference

Let us assume that we have m training (i.e., labeled) patterns $\vec{x}_p = (x_{p1},\ ...,\ x_{pn})$, $p = 1,\ 2,\ ...,\ m$ from M classes in an n-dimensional pattern space where x_{pi} is the attribute value of the pth pattern for the ith attribute ($i = 1,\ ...,\ n$). For the simplicity of explanation, we assume that all the attribute values have already been normalized into real numbers in the unit interval [0, 1]. Thus the pattern space of our classification problem is an n-dimensional unit-hypercube $[0, 1]^n$.

For our n-dimensional pattern classification problem, we use fuzzy rules of the following type:

$$R_q:\ \text{If } x_1 \text{ is } A_{q1} \text{ and } ... \text{ and } x_n \text{ is } A_{qn} \text{ then Class } C_q \text{ with } CF_q, \quad (1)$$

where R_q is the label of the qth fuzzy rule, $\vec{x} = (x_1,\ ...,\ x_n)$ is an n-dimensional pattern vector, A_{qi} is an antecedent fuzzy set ($i = 1,\ ...,\ n$), C_q is a class label, and CF_q is a rule weight. We denote the antecedent fuzzy sets of R_q as a fuzzy vector $\vec{A}_q = (A_{q1},\ A_{q2},\ ...,\ A_{qn})$.

Fourteen fuzzy sets are initially considered in four fuzzy partitions with different granularities. Figure 1 depicts these partitions. In addition to those 14 fuzzy sets, we also use the domain interval [0, 1] itself as an antecedent fuzzy set in order to represent a *don't care* condition.

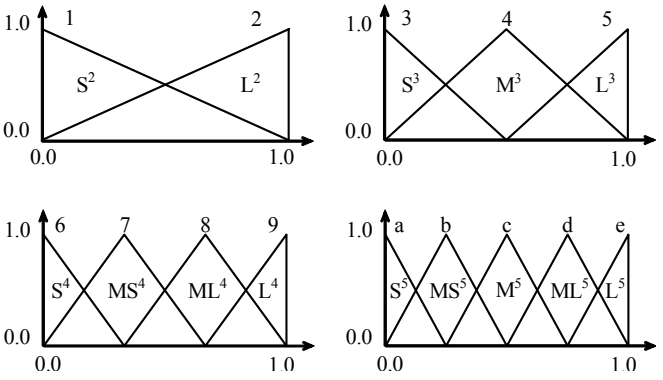

Fig. 1. The fourteen antecedent fuzzy sets considered.

Let S be a set of fuzzy rules of the form in (1). When an input pattern $\vec{x}_p$ is to be classified by S, first we calculate the compatibility grade of $\vec{x}_p$ with the antecedent part $\vec{A}_q = (A_{q1},\ A_{q2},\ ...,\ A_{qn})$ of each fuzzy rule R_q in S using the product operation as,

$$\mu_{\vec{A}_q}(\vec{x}_p) = \mu_{A_{q1}}(x_{p1}) \cdot ... \cdot \mu_{A_{qn}}(x_{pn}), \quad (2)$$

where $\mu_{A_{qi}}(\cdot)$ is the membership function of the antecedent fuzzy set A_{qi}. Then a single winner rule R_w is identified using the compatibility grade and the rule weight of each fuzzy rule as

$$\mu_{\vec{A}_w}(\vec{x}_p) \cdot CF_w = \max\{\mu_{\vec{A}_q}(\vec{x}_p) \cdot CF_q \mid R_q \in S\}. \quad (3)$$

The input pattern $\vec{x}_p$ is classified as the consequent class C_w of the winner rule R_w. When multiple fuzzy rules with different consequent classes have the same maximum value in (3), the classification of $\vec{x}_p$ is rejected. If there is no compatible fuzzy rule with $\vec{x}_p$, its classification is also rejected.

III. An Algorithm for Generating Single Granularity-Based Fuzzy Classification Rules

As we have already explained, multiobjective genetic fuzzy rule selection has been based on a previously fixed granularity [10], [11] (five linguistic terms in all the attributes) or multiple granularities [12]. Based on this last approach [12], in this section we propose a mechanism to generate single granularity-based fuzzy classification rules, a nearer to the interpretability approach. The proposed procedure is as follows:

- *Step 1:* Rule extraction with multiple granularities.
- *Step 2:* Specification of single granularity for each attribute based on the extracted rules.
- *Step 3:* Rule extraction with selected single granularities.
- *Step 4:* Multiobjective genetic fuzzy rule selection.

The original multiple granularities based procedure [12] is composed of Steps 1 and 4. Steps 2 and 3 are additional procedures. In Step 1, we extract a fixed short number of rules for each class based on well-known data mining rule evaluation measures [2] and multiple granularities. In Step 2, we select a single granularity for each attribute based on the extracted rules. Then, we extract the final set of candidate rules for each class by using the selected single granularities in Step 3. Step 4 is the same as the original one to perform multiobjective genetic fuzzy rule selection. The next subsections present detailed explanations of these steps.

A. Rule Extraction with Multiple Granularities (Step 1)

Since 14 antecedent fuzzy sets in Figure 1 and an additional *don't care* fuzzy set [0, 1] are used for each attribute of the n-dimensional classification problem, the total number of possible fuzzy rules is 15^n. Among these possible rules, we examine only short fuzzy rules with a small number of antecedent conditions (i.e., short fuzzy rules with many *don't care* conditions) to generate an initial set of candidate rules. In this work, we specify the maximum number of antecedent conditions as three for datasets with less than 30 attributes and two for datasets with more than or equal to 30 attributes.

The consequent class C_q and the rule weight CF_q of each fuzzy rule R_q are specified from training patterns compatible with its antecedent part $\vec{A}_q = (A_{q1},\ A_{q2},\ ...,\ A_{qn})$ in the

following heuristic manner [13]. First the confidence of each class for the antecedent part $\vec{A}_q$ is calculated as:

$$c(\mathbf{A}_q \Rightarrow \text{Class } h) = \frac{\sum_{\mathbf{x}_p \in \text{ Class } h} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}{\sum_{p=1}^{m} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}, \quad h = 1, \ ..., \ M. \tag{4}$$

It should be noted that "$\mathbf{A}_q \Rightarrow$ Class h" means the fuzzy rule with the antecedent part $\vec{A}_q$ and the consequent class h. Then the consequent class C_q is specified by identifying the class with the maximum confidence:

$$c(\mathbf{A}_q \Rightarrow \text{Class } C_q) = \max_{h=1, \ 2, \ ..., \ M} \{c(\mathbf{A}_q \Rightarrow \text{Class } h)\}. \tag{5}$$

In this manner, we generate the fuzzy rule R_q (i.e., $\mathbf{A}_q \Rightarrow$ Class C_q) with the antecedent part $\vec{A}_q$ and the consequent class C_q. We do not generate any fuzzy rules with the antecedent part $\vec{A}_q$ if there is no compatible training pattern with $\vec{A}_q$.

The rule weight CF_q of each fuzzy rule R_q has a large effect on the performance of fuzzy rule-based classifiers. We use the following specification of CF_q because good results were reported in the literature [14]:

$$CF_q = c(\mathbf{A}_q \Rightarrow \text{Class}\, C_q) - \sum_{\substack{h=1 \\ h \neq C_q}}^{M} c(\mathbf{A}_q \Rightarrow \text{Class}\, h). \tag{6}$$

We do not use the fuzzy rule R_q as a candidate rule if the rule weight CF_q is not positive (i.e., if its confidence is not larger than 0.5).

In the above-mentioned heuristic manner, we can generate a large number of short fuzzy rules as candidate rules in multiobjective fuzzy rule selection (some of them with not interesting properties). In order to directly focus on the most interesting rules, a prescreening procedure is applied to decrease the number of candidate rules. This prescreening procedure is based on well-known rule evaluation measures in the field of data mining [2]: *support* and *confidence*.

For prescreening candidate rules, we use two threshold values: the minimum support and the minimum confidence. We exclude fuzzy rules that do not satisfy these two threshold values. Among short fuzzy rules satisfying these two threshold values, we choose a prespecified number of candidate rules for each class. As a rule evaluation criterion, we use the product of the support $s(R_q)$ and the confidence $c(R_q)$. That is, we choose a prespecified number of the best candidate rules for each class with respect to product $p(R_q) = s(R_q) \cdot c(R_q)$.

B. Single Granularity Specification and Rule Extraction (Steps 2 and 3)

Once a set of candidate rules is obtained based on multiple granularities (Step 1), the original approach [12] goes to Step 4 in order to apply multiobjective fuzzy rule selection. However, there is useful information in the extracted rules that could be used to specify an appropriate single granularity for each attribute. Frequency of the employed granularities in the extracted rules (weighted by the corresponding rule importance) can be used to fix the most promising granularities. For each attribute i ($i = 1, \ ..., \ n$), we specify the granularity with the highest sum of importance of the rules considering such granularity in the corresponding attribute:

$$Gr(i) = \underset{g=2, \ ..., \ 5}{\text{argmax}} \left\{ \sum_{Gran(\mathbf{A}_{qi})=g} \text{Imp}(\mathbf{R}_q) \right\}, \tag{7}$$

where $Gran(\mathbf{A}_{qi})$ is the granularity of the partition containing the fuzzy set used in attribute i of rule R_q and $\text{Imp}(R_q)$ is a criterion associated to the importance of the rule in the sum. Many criteria can be considered involving different specification mechanisms:

- Frequency: $\text{Imp}(R_q) = 1, \forall q$.
- Confidence: $\text{Imp}(R_q) = c(R_q), \forall q$.
- Weight: $\text{Imp}(R_q) = CF_q, \forall q$.
- Support: $\text{Imp}(R_q) = s(R_q), \forall q$.
- Product: $\text{Imp}(R_q) = p(R_q), \forall q$.

However, the first three criteria are not recommended since they usually provoke overfitting. We will study the last two criteria as a way to extract more general rules instead of very specific ones, which helps to the generalization ability. In the same way, in order to preferably take into account more general rules we examine two approaches named, 1-ALL approach and 1-2-3 approach, with the two basic criteria (i.e., product and support). Both approaches give priority to granularities in the rules with a single condition, i.e., Equation (7) is applied by only considering size one rules if possible. The difference is only when there is no rule with a single condition in the corresponding attribute. Let us consider the product criterion and the next six rules, where g^i represents any fuzzy set of a partition with granularity i,

R_1 : If x_1 is g^2 and x_2 is g^4 and x_3 is g^3 then Class 1, $p(R_1) : 0.4$.
R_2 : If x_1 is g^4 then Class 2, $p(R_2) : 0.8$.
R_3 : If x_2 is g^3 then Class 2, $p(R_3) : 0.3$.
R_4 : If x_2 is g^2 then Class 1, $p(R_4) : 0.8$.
R_5 : If x_2 is g^3 and x_3 is g^4 then Class 1, $p(R_5) : 0.6$.
R_6 : If x_1 is g^2 and x_2 is g^2 and x_3 is g^3 then Class 1, $p(R_6) : 0.3$.

When we specify a granularity for the first attribute, we first check rule(s) with a single condition related to the first attribute by both approaches (1-ALL and 1-2-3). Since rule R_2 is the only rule in this situation, we select granularity 4 for the first attribute. Next, in the same manner, we can find two rules: R_3 and R_4 for the second attribute. We select Granularity 2 for the second attribute because of the high product value by both approaches. Finally, we select a granularity for the third attribute but there is no rule with a single condition. In 1-ALL approach, we specify a single granularity from all the rules including the third attribute

independently of the number of conditions they have (rules R_1, R_5 and R_6). The sum of product values for granularity 3 is 0.7 and 0.6 for granularity 4. From this comparison, we select granularity 3 for the third attribute. On the other hand, in 1-2-3 approach, we give priority to the rules with a smaller number of conditions (two conditions). That is, we select granularity 4 for the third attribute (if there are no rules with two conditions then those with three are considered).

In Step 2, we select a single granularity for each attribute. Four different mechanisms have been defined: Support/1-ALL, Product/1-ALL, Support/1-2-3 and Product/1-2-3. Finally, in Step 3, we apply again the candidate rule extraction procedure explained in Step 1 by only using the specified single granularities for each attribute.

C. Multiobjective Fuzzy Rule Selection (Step 4)

Let us assume that we have N candidate rules (i.e., N/M candidate rules for each of M classes). Any subset S of the N candidate rules can be represented by a binary string of length N: $S = s_1 s_2 ... s_N$ where $s_j = 1$ and $s_j = 0$ mean the inclusion and the exclusion of the jth candidate rule R_j in the subset S, respectively (j = 1, ..., N). Such a binary string S is used as an individual in an EMO algorithm for multiobjective fuzzy rule selection.

It should be noted that S can be viewed as a fuzzy rule-based classifier. Each fuzzy rule-based classifier S is evaluated by the next three objectives:

- $f_1(S)$: the number of correctly classified training patterns.
- $f_2(S)$: the number of selected fuzzy rules.
- $f_3(S)$: the total number of antecedent conditions.

That is, our multiobjective fuzzy rule selection problem is written as:

$$\text{Maximize } f_1(S), \text{ and minimize } f_2(S) \text{ and } f_3(S). \quad (8)$$

We use NSGA-II of Deb et al. [8] to search for non-dominated fuzzy rule-based classifiers with respect to these three objectives, uniform crossover and bit-flip mutation. The execution of NSGA-II was terminated at the prespecified number of generations.

In order to efficiently decrease the number of fuzzy rules in each rule set S, two heuristic techniques are used. One is biased mutation where a larger mutation probability is assigned to the mutation from 1 to 0 than that from 0 to 1. The other is the removal of unnecessary fuzzy rules. Since we use the single winner-based scheme in (3) for classifying each training pattern by a fuzzy rule-based classifier S, some fuzzy rules in S may classify no training patterns. We can remove those unnecessary fuzzy rules from S without changing any classification results by S (i.e., without changing the first objective $f_1(S)$). This heuristic procedure can be viewed as a kind of local search since $f_2(S)$ and $f_3(S)$ are improved without deteriorating $f_1(S)$.

IV. Experiments

In order to examine the effects of the granularity specification mechanisms proposed in this paper, we have selected 15 datasets from the UCI repository [3]. Table I summarizes the properties of these datasets. It shows, for each dataset, the number of patterns, the number of attributes and the number of classes. In the case of presenting missing values (Aut, Cleveland and Dermatology) we have removed the instances with any missing value before partitioning.

TABLE I
DATASETS CONSIDERED FOR COMPARISONS

Name	Patterns	Attributes	Classes
Appendicitis	106	7	2
Aut	159	25	4
Cleveland	297	13	5
Dermatology	358	34	6
Glass	214	9	6
Haberman	306	3	2
Hayes-roth	132	4	3
Iris	150	4	3
Newthyroid	215	5	3
Pasture	36	22	3
Pima	768	8	2
Saheart	462	9	2
Tae	151	5	3
Vehicle	846	18	4
Wine	178	13	3

In order to analyze the performance of the single granularity specification, we compare the original approach (All Granularities) with the four mechanisms proposed to specify single granularities (Support/1-ALL, Product/1-ALL, Support/1-2-3 and Product/1-2-3). The parameter settings for all the considered approaches are as follows (same conditions in all the cases):

- The number of fuzzy rules for each class: 300.
- Optimizer: NSGA-II.
- Population size: 200.
- The number of generations: 5000.
- Crossover probability: 0.9 (Uniform crossover).
- Mutation probability: 0.05 (from 1 to 0), $1/L$ (from 0 to 1, where L is the string length).

We consider a *10-fold cross-validation model*, i.e., 10 random partitions of data each with 10%, and the combination of 9 of them (90%) as training and the remaining one as test. For each one of the 10 data partitions, the studied methods have been run 3 times, showing for each problem the averaged results of a total of 30 runs (10fcv x 3 different random seeds). Since these methods present a multi-objective nature, the averaged values are calculated considering the most accurate solution from each Pareto front obtained (the one with the highest classification rate in training). Our main aim following this approach is to obtain more reliable information at least in this part of the Pareto front, which in any case is comprised by quite simple models.

Table II shows the averaged number of rules/conditions (#R / #C) and classification percentages in training (Tr.) and test (Ts.) of the most accurate classifier from each of the obtained Pareto fronts. The overall mean values for each method are in the last row. Taking into account these results we can highlight the following facts:

TABLE II

RESULTS OBTAINED BY THE STUDIED METHODS (MOST ACCURATE)

	All Granularities				Support/1-ALL				Product/1-ALL				Support/1-2-3				Product/1-2-3			
Datasets	#R	#C	Tr.	Ts.	#R	#C	Tr.	Ts.	#R	#C	Tr.	Ts.	#R	#C	Tr.	Ts.	#R	#C	Tr.	Ts.
Appendicitis	2.37	3.7	91.86	87.91	3.40	6.9	93.22	88.21	3.40	7.0	93.29	88.21	3.40	6.9	93.22	88.21	3.40	7.0	93.29	88.21
Aut	9.87	24.1	81.59	66.52	11.60	29.1	86.49	69.33	11.67	29.3	87.19	69.91	12.20	31.0	87.30	68.38	12.33	30.5	87.09	68.56
Cleveland	20.17	56.1	73.11	55.11	25.07	68.2	70.37	51.08	28.67	77.8	77.28	52.83	24.80	67.8	69.06	52.28	28.93	78.9	76.72	53.30
Dermatology	11.40	19.3	99.07	94.12	13.10	22.7	99.62	94.93	13.53	23.5	99.51	93.26	13.10	22.7	99.62	94.93	13.53	23.5	99.51	93.26
Glass	12.63	32.4	78.65	60.48	16.97	39.0	80.74	66.07	19.30	44.7	83.85	69.96	17.83	41.2	81.23	64.42	19.40	44.7	83.97	68.29
Haberman	6.50	13.9	79.46	71.89	2.50	4.9	74.44	73.19	3.00	6.0	74.70	73.19	2.50	4.9	74.44	73.19	3.00	6.0	74.70	73.19
Hayes-roth	9.17	15.3	90.88	78.03	10.80	16.7	90.79	79.14	10.83	16.9	90.91	79.14	10.80	16.7	90.79	79.14	10.83	16.9	90.91	79.14
Iris	4.03	6.8	99.11	95.11	5.23	7.5	98.30	94.67	5.23	7.5	98.30	95.33	5.23	7.5	98.30	94.67	5.23	7.5	98.30	95.33
Newthyroid	5.37	9.2	96.19	91.78	7.60	15.5	97.71	94.83	7.37	15.7	97.59	93.01	7.60	15.5	97.71	94.83	7.37	15.7	97.59	93.01
Pasture	3.70	5.9	98.05	75.83	4.27	7.7	99.69	73.61	4.43	8.1	100.00	73.61	4.27	7.5	99.58	73.61	4.63	8.4	99.79	77.50
Pima	6.63	14.3	77.80	74.92	10.63	25.9	79.04	73.79	10.63	25.9	79.04	73.79	10.63	25.9	79.04	73.79	10.63	25.9	79.04	73.79
Saheart	5.97	12.8	76.70	71.14	12.33	31.8	79.00	71.22	12.33	31.8	79.00	71.22	12.33	31.8	79.00	71.22	12.33	31.8	79.00	71.22
Tae	7.77	18.9	66.55	54.57	10.77	24.5	69.36	59.22	11.33	25.4	69.88	59.24	9.80	22.1	67.99	55.90	8.53	18.8	65.98	57.03
Vehicle	13.77	35.8	69.34	62.81	15.73	43.1	71.11	66.16	15.60	43.1	71.07	66.20	16.50	44.3	70.69	66.71	16.63	45.1	70.80	66.51
Wine	3.90	8.2	100.00	96.08	6.00	11.7	99.98	93.26	6.37	12.1	99.92	95.11	6.07	11.8	99.98	92.67	6.37	12.1	99.92	94.52
Mean	**8.22**	**18.5**	85.22	75.75	10.40	23.7	85.99	76.58	10.91	25.0	**86.77**	**76.93**	10.47	23.8	85.86	76.26	10.88	24.9	86.44	76.86

TABLE III

RESULTS OBTAINED BY THE STUDIED METHODS (EQUIVALENT COMPLEXITY)

	Support 1-ALL				Product 1-ALL				Support 1-2-3				Product 1-2-3			
Datasets	#R	#C	Tr.	Ts.	#R	#C	Tr.	Ts.	#R	#C	Tr.	Ts.	#R	#C	Tr.	Ts.
Appendicitis	2.23	4.1	92.17	88.82	2.17	4.1	92.24	88.82	2.23	4.1	92.17	88.82	2.17	4.1	92.24	88.82
Aut	9.67	24.2	85.11	68.92	9.73	24.4	85.84	68.94	10.17	25.9	85.95	68.01	9.80	24.4	85.20	69.73
Cleveland	20.07	53.2	68.51	51.51	20.20	53.4	74.14	53.48	19.83	52.9	67.22	52.39	20.07	53.4	73.29	53.28
Dermatology	11.23	19.3	99.02	94.65	11.70	20.1	98.90	93.08	11.23	19.3	99.02	94.65	11.70	20.1	98.90	93.08
Glass	12.60	27.7	78.35	66.82	12.57	28.1	80.15	69.90	12.43	27.2	78.11	66.14	12.73	28.2	80.27	68.54
Haberman	2.50	4.9	74.44	73.19	3.00	6.0	74.70	73.19	2.50	4.9	74.44	73.19	3.00	6.0	74.70	73.19
Hayes-roth	9.53	13.1	89.95	80.09	9.57	13.4	90.07	80.09	9.53	13.1	89.95	80.09	9.57	13.4	90.07	80.09
Iris	4.23	5.3	97.56	94.44	4.23	5.3	97.56	94.44	4.23	5.3	97.56	94.44	4.23	5.3	97.56	94.44
Newthyroid	5.30	10.1	96.21	94.42	5.63	11.2	96.54	92.23	5.30	10.1	96.21	94.42	5.63	11.2	96.54	92.23
Pasture	3.83	6.5	97.22	75.00	3.93	6.4	96.90	73.89	3.67	6.3	96.80	69.72	3.93	6.3	96.60	73.61
Pima	6.87	16.0	78.36	73.78	6.87	16.0	78.36	73.78	6.87	16.0	78.36	73.78	6.87	16.0	78.36	73.78
Saheart	6.20	14.9	76.71	70.49	6.20	14.9	76.71	70.49	6.20	14.9	76.71	70.49	6.20	14.9	76.71	70.49
Tae	7.70	15.4	64.82	57.93	7.67	15.1	64.73	57.04	7.83	16.1	65.61	57.01	7.77	16.3	65.15	58.36
Vehicle	13.87	37.6	70.83	66.27	13.77	37.7	70.80	66.08	14.00	37.0	70.32	66.63	14.07	37.6	70.43	66.43
Wine	3.87	7.0	96.71	90.64	3.93	5.3	96.17	90.99	3.70	7.1	96.55	89.48	3.93	5.4	96.19	91.72
Mean	**7.98**	**17.3**	84.40	76.47	8.08	17.4	**84.92**	76.43	**7.98**	17.4	84.33	75.95	8.11	17.5	84.81	**76.52**

- In general, all the proposed approaches present better classification rates than the original one, obtaining the best overall result in the case of Product/1-ALL. All of them get better test performance in 10 of the 15 datasets (11 in the case of Product/1-2-3), with approximately two more rules with respect to the original model considering multiple granularities.
- Product based approaches present slightly better results than support based ones.

In Table III, we show the average results of the proposed approaches by considering the closest solution in number of rules (equivalent complexity) to the most accurate solution from All Granularities approach for each of the 30 Pareto fronts. Even though that it seems preferable to ensure a single granularity than to obtain a simpler solution, with approximately two less rules, we can observe in this table that very similar results are obtained when equivalent complexities are considered. This demonstrates that by fixing an appropriate single granularity at least equivalent results can be obtained from the point of view of the accuracy and the complexity. Further, it is highly preferable to avoid multiple granularities in terms of the global interpretability of the obtained models.

Figure 2 plots the average Pareto fronts on training and test sets, composed by the average values of the different solutions of each of the thirty Pareto fronts (from the most accurate solution to the simplest one). This method represents an extension of the idea of analyzing the most accurate solutions in the Pareto fronts (first, second, etc.) presented in [9] where the search is focused on the most accurate solutions only. We consider the next most accurate solution of each of the 30 fronts to compute a new averaged solution, until no more solutions remain in any of these fronts. As we can see in this example, by determining single granularities the test accuracy can be improved in most of the solutions obtained. As an example, two rule sets obtained by All Granularities and Product/1-ALL are depicted in Figure 3.

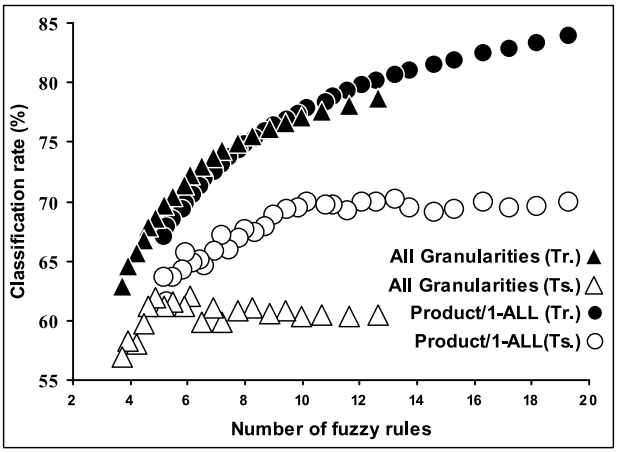

Fig. 2. Average Pareto fronts (training and test) for Glass dataset with All Granularities, Product/1-ALL and Product/1-2-3.

V. Concluding Remarks

In this work, we have proposed a method to generate single granularity-based fuzzy classification rules for multiobjective genetic fuzzy rule selection. After extracting a prespecified number of fuzzy rules with multiple granularities, a single granularity has been specified for each attribute individually according to the frequency of employed partitions and the importance of the multiple granularity-based extracted rules. Then multiobjective genetic fuzzy rule selection is applied in order to obtain a set of non-dominated solutions with a good trade-off between complexity and accuracy.

Following the same main idea, four different mechanisms (Support/1-ALL, Product/1-ALL, Support/1-2-3 and Product/1-2-3) have been proposed and compared with the original approach on a set of 15 well-known datasets. As well as the interpretability improvement that the use of a fixed single granularity involves, the results show that the performance of the obtained classifiers can be even improved by avoiding multiple granularities. Further, we can conclude that the most promising mechanisms are those based on Product criterion with no clear differences between 1-ALL and 1-2-3 based approaches.

All Granularities (training: 79.17, test: 63.64)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	class
R_1	DC	DC	DC		DC	DC		DC	DC	1 (0.03)
R_2	DC	DC			DC	DC	DC		DC	1 (0.12)
R_3	DC	DC	DC	DC	DC	DC	DC	DC		2 (0.04)
R_4		DC	DC	DC	DC		DC	DC		2 (0.08)
R_5		DC			DC	DC	DC	DC	DC	2 (0.21)
R_6	DC	DC	DC	DC		DC	DC		DC	2 (0.19)
R_7	DC	DC	DC			DC	DC	DC		3 (0.98)
R_8	DC			DC	DC	DC	DC	DC		3 (0.51)
R_9				DC	DC	DC	DC	DC	DC	4 (0.39)
R_{10}	DC			DC		DC	DC	DC	DC	5 (0.10)
R_{11}	DC		DC	DC	DC	DC	DC		DC	5 (0.85)
R_{12}	DC	DC			DC	DC		DC	DC	6 (0.90)
R_{13}	DC	DC		DC	DC	DC	DC		DC	6 (0.99)

Product/1-ALL (training: 76.53, test: 72.22)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	class
R_1		DC		DC	DC	DC	DC	DC	DC	1 (0.30)
R_2		DC	DC	DC	DC		DC	DC		1 (0.60)
R_3		DC	DC	DC	DC	DC	DC	DC	DC	2 (0.56)
R_4	DC	DC			DC	DC	DC	DC	DC	2 (0.33)
R_5	DC	DC			DC	DC	DC	DC	DC	2 (0.17)
R_6	DC	DC		DC	DC	DC	DC	DC		2 (0.61)
R_7			DC	DC	DC	DC	DC	DC		3 (0.79)
R_8	DC		DC		DC	DC		DC	DC	4 (0.31)
R_9	DC	DC	DC		DC	DC	DC	DC	DC	4 (0.64)
R_{10}				DC	DC	DC	DC	DC	DC	5 (0.23)
R_{11}		DC	DC	DC		DC	DC	DC	DC	6 (0.17)
R_{12}			DC	DC	DC	DC	DC		DC	6 (0.92)
R_{13}	DC	DC		DC	DC	DC		DC	DC	6 (0.15)

Fig. 3. Two rule sets with multiple (#R=13, #C=33) and Product/1-ALL single (#R=13, #C=29) granularity for Glass dataset.

References

[1] R. Alcalá, J. R. Cano, O. Cordón, F. Herrera, P. Villar and I. Zwir, "Linguistic modeling with hierarchical systems of weighted linguistic rules," *Approximate Reasoning* vol. 32, no. 2-3, pp. 187–215, 2003.

[2] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A.I. Verkamo, "Fast discovery of association rules. In: Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R (Ed.)," *Advances in Knowledge Discovery and Data Mining*, AAAI (Menlo Park, 1996) pp. 307–328.

[3] A. Asuncion and D.J. Newman. UCI machine learning repository, 2007. URL http://www.ics.uci.edu/mlearn/MLRepository.html.

[4] J. Casillas, O. Cordón, F. Herrera, and L. Magdalena (Eds.), "Interpretability issues in fuzzy modeling," *Studies in Fuzziness and Soft Computing* 129 (Springer, Heidelberg, Germany, 2003).

[5] M. Cococcioni, P. Ducange, B. Lazzerini, and F. Marcelloni, "A Pareto-based multi-objective evolutionary approach to the identification of mamdani fuzzy systems," Soft Computing vol. 11, pp. 1013–1031, 2007.

[6] C.A. Coello, D.A. Van Veldhuizen, and G.B. Lamont, "Evolutionary Algorithms for Solving Multi-Objective Problems", Kluwer Academic Publishers (2002).

[7] K. Deb, "Multi-objective optimization using evolutionary algorithms," John Wiley & Sons, Chichester (UK, 2001).

[8] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II, " *IEEE Transactions on Evolutionary Computation* vol. 6, no. 2, pp. 182–197, 2002.

[9] M.J. Gacto, R. Alcalá, and F. Herrera, "Adaptation and Application of Multi-Objective Evolutionary Algorithms for Rule Reduction and Parameter Tuning of Fuzzy Rule-Based Systems," *Soft Computing* vol. 13, no. 5, pp. 419–436, 2009.

[10] H. Ishibuchi, T. Murata, and I.B. Türksen, "Selecting linguistic classification rules by two-objective genetic algorithms," *Proc. Proceedings of the SMC* (1995) pp. 1410–1415

[11] H. Ishibuchi, T. Murata, and I.B. Turksen, "Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems," *Fuzzy Sets and Systems* vol. 89, no. 2, pp. 135–150, 1997.

[12] H. Ishibuchi, and T. Yamamoto, "Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining," *Fuzzy Sets and Systems* vol. 141, no. 1, pp. 59–88, 2004.

[13] H. Ishibuchi, T. Nakashima, and M. Nii, "Classification and modeling with linguistic information granules: Advanced approaches to linguistic data mining," Springer (Berlin, 2004).

[14] H. Ishibuchi, and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," *IEEE Transactions on Fuzzy Systems* vol. 13, pp. 428–435, 2005.